

Vishesh Chahar

Phone: (+91)-9501006533

Email: vishesh.chahar01@gmail.com

LinkedIn: www.linkedin.com/in/visheshchahar

GitHub: www.github.com/Vishesh-Chahar

LeetCode: www.leetcode.com/vishesh_chahar

BIO

Data Scientist specializing in **Machine Learning pipelines**, **Natural Language Processing**, and **LLM fine-tuning**. Proficient in **Python**, leveraging **CUDA-enabled PyTorch** for accelerated model training. Experienced in **ELT pipeline optimization** using **Kafka** and **PySpark**, and building scalable **REST APIs** with **Flask**. Skilled in **CI/CD** with **Docker** for containerization and **Git** for version control.

EDUCATION

Thapar Institute of Engineering and Technology, Patiala, Punjab

Bachelor of Engineering in Computer Engineering

Oct 2020 - Jun 2024

Bhavan Vidyalaya, Chandigarh

Senior Secondary School

Mar 2018 - Mar 2020

St. John's High School, Chandigarh

Secondary School

Mar 2006 - Mar 2018

EXPERIENCE

Isourse Technologies

Jun 2024 - Dec 2024

AI/ML RND Team

Gen-BI

[Llama 3.1, PostgreSQL, Flask, Python]

- Led a cross-functional RND team to develop a novel BI module with **Llama 3.1** reducing skill dependence and manual reporting by **20+ hours/month**
- Crafted and maintained secure, scalable APIs using **Flask** supporting over 1000 concurrent users, leading to reduced application latency

Data Warehousing

[Phi-4, Apache Kafka, PySpark, Docker, K-Means]

- Lowered data storage and processing costs by **30%** with a hybrid Data Warehousing solution integrated with **Phi-4**, **K-Means**
- Improved pipeline efficiency by **42% (74s to 43s)** and transformation speed by **34% (37s to 24s)** by using **Kafka**-based data streaming and optimizing **Spark Executors**.

Computer Vision

[YOLO, Pytorch, CNN, LSTM]

- Accomplished **83% accuracy on a custom CNN for handwritten OCR** on multi-digit numbers with a model size of 6.7 MB
- Attained **66.4% accuracy on object detection task** using **YOLO** for document digitization on self-annotated dataset

Wipro Limited

Jan 2024 - Jun 2024

Data Science Intern

[BERTForQA, Llaama 3, Scikit-Learn, Pandas, BeautifulSoup]

- Decreased data preprocessing delays by **13%** by implementing automation script for data extraction using **BeautifulSoup**
- Minimized **manual work hours by upto 30 hours/month** by automation of data extraction and processing using **Pandas**
- Increased system reliability by utilizing **AWS Sagemaker** to containerize pipelines
- Achieved **r2 score of 0.92** with **regularized in-house regression model** for prediction tasks

PROJECTS

Headliner	Project Link	Python, CUDA, NLP, flan-t5	Diagnosis Pal	Project Link	Python, ML, CategoricalNB
		<ul style="list-style-type: none">Fine-tuned 27.36% of parameters of flan-t5-large with LoRA for headline generation from 236 articles with over 600 tokens each.Attained a 150% decrease in model training time by implementing GPU computations using CUDA.			<ul style="list-style-type: none">Strategized data preprocessing techniques for disease classification using CategoricalNB from symptom dataEnhanced predictive precision, achieving a 84% MAP@K score, and improved to 92% MAP@K with hyperparameter adjustment.
QA	Project Link	NLP, huggingface, BERTForQA	Team Fateh	Project Link	Powertrain, EV, DAQ
		<ul style="list-style-type: none">Lessened manual data entry by 50% by manufacturing custom solutions using Llama-2 and BertForQA to enable focus on strategic tasks.			<ul style="list-style-type: none">Championed 21 trophies and INR 4,10,000 over 4 years.Collaborated with a team of 60 engineers to 7 international competitions.

SKILLS

- Languages:** Python, R, C++, C#, SQL, Bash, Shell
- AI/ML:** PyTorch, Keras, TensorFlow, CUDA, Scikit-learn
- NLP:** LLaMA, BERT, FLAN-T5, NLTK, Hugging Face
- CV:** YOLO, CNN, LSTM
- CI/CD:** Flask, Git, Docker, Postman
- Data Engineering:** PostgreSQL, PySpark, Kafka, Pandas